DOCUMENT RESUME

ED 064 336 TM 001 501

AUTHOR Nettleton, Aileen L.
TITLE Taming the Standardized Testing Program.
PUB DATE Apr 72
NOTE 13p.; Symposium paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Achievement Tests; City Wide Programs; *Mathematics; *Reading Tests; *Student Testing; *Testing Programs
IDENTIFIERS *Madison; Wisconsin

ABSTRACT

The city-wide standardized testing program of Madison Public Schools was reviewed by a committee of a cross-section of school system educators as part of a total effort to design a testing program more sensitive to the needs of the system. As a result, standardized testing was reduced to reading (grades 1, 2, 3, 4, 5, and 8) and mathematics (grade 5). Levels of administration were determined by the importance of measuring reading progress in elementary grades and the value of achievement level indicators in transition between elementary, middle and high school. Under this plan, standardized tests are intended to provide normative data to compare the school system with others, to evaluate educational programs within the system, and to give an indication of student achievement ranking. (Author/DB)

ED 064336

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

TAMING THE STANDARDIZED TESTING PROGRAM

Aileen L. Nettleton

Reading Consultant
MADISON PUBLIC SCHOOLS

1972

Presented as part of a Symposium entitled "The Madison Plan: A New Approach to System-wide Testing" at the 1972 convention of the American Educational Research Association.

TM 001 501

Introduction

Extensive standardized testing has characterized the city-wide testing program in Madison for some years. Teachers and administrators alike were beginning to question the value of such a massive, time-consuming and costly operation in respect to benefits accrued. A committee of teachers, counselors, psychologists and administrators under the direction of Dr. T. Anne Cleary, University of Wisconsin psychometrician, was set up to study current needs for test information in the school system and design a testing program to meet those needs. From a survey needs of Madison educators for information about students, three different types of testing needs emerged: standardized, criterion-referenced or curriculum related, and affective. This paper will trace the development of the standardized test program.

Designing the Standardized Testing Program

The results of the survey of testing needs of Madison educators showed that the current standardized testing program might purport to answer only two of the twelve questions rated highest: "What is the child's capacity for learning?" and "What general achievement level does the child have in reading?" A major question which standardized tests are intended to answer, "How does the child's academic achievement

compare with that of other children in the nation?" was rated low by most segments of the educational community.

These results, plus the concern of many committee members that standardized testing was too extensive and provided little meaningful data, showed that one aspect of a testing program to which the committee must address itself was standardized testing. Several basic questions emerged from committee members: Why administer an entire battery of achievement tests if only reading scores are needed? Who uses the normative information gained from standardized tests? Should we completely eliminate standardized testing and replace it with curriculum-related testing to meet teachers' needs for diagnostic information?

A meeting with the Superintendent of Schools revealed that standardized test results are, indeed, significant to him as he is called upon to be accountable to the public for educational results and for making program and budget decisions to improve education throughout the city. Normative data provided by standardized tests is needed as a yardstick to check the status of local educational outcomes against education in the nation as a whole. Teachers on the committee indicated that if standardized test results were reported back immediately following testing, the information might assist them in student grouping and placement decisions. Guidance counselors expressed a need for standardized test scores to monitor achievement patterns of individual students and to assist students in making educational and vocational choices. In addition to finding that certain groups did feel a need for nor-

mative, standardized testing, other members of the committee recognized that until curriculum-related tests were available for city-wide use, they preferred limited standardized test data rather than none for general academic program evaluation.

With the need for some type of standardized testing program evident, the question was how much and what kinds would satisfy needs? To equip the committee for making these decisions, time was devoted to instruction in several basic measurement concepts. Construct validity was examined in respect to the issue of the difference between the construct "intelligence" and the measure "group intelligence test." Correlation among standardized tests being administered as part of the city-wide testing program were studied to determine if different behaviors were actually being measured through the battery of tests.

The correlation between the group IQ and reading tests, computed by the test publishers, ranged from .60 to .84[1]. This supported the feelings expressed by some members that group intelligence test scores depended on student reading ability. An examination of intercorrelations among the achievement tests in the battery being administered at several grade levels showed correlations of .69 to .82 among the tests (Vocabulary, Reading, Language Skills, Work-Study Skills, Arithmetic Skills) and .88 to .92 between Reading and Composite scores.[2] Thus, the bulk of information being obtained from an extensive testing program was redundant. Scores from the reading test would be an adequate indicator of achievement in language skills, work study skills, and, in many cases, even mathematics.

The concept of item and student sampling was also introduced as a procedure for answering questions about instructional programs more efficiently. Subtests or sets of items could be randomly administered to samples of students and still allow for the same inferences to be made about individual school or system programs as from administering an entire test or to all students. A sampling design would not provide data on individual student achievement, however, and it was the consensus of committee members that until criterion-referenced tests become available, all students at designated levels should be tested.

Recommendations

Consideration of these factors--actual needs for certain standardized test information and the limited amount of actual information gained through wide testing--led to a decision by the Nucleus Testing Committee to recommend a reduction of testing to administration of only a reading achievement test at six grade levels and a mathematics achievement test at one grade level:

Grade 1	Fall	Reading Readiness
Grade 2	Fall	Reading Achievement
Grade 3	Fall	Reading Achievement
Grade 4	Fall	Reading Achievement
Grade 5	Winter	Reading Achievement Mathematics Achievement
Grade 8	Winter	Reading Achievement

These grade levels were selected as critical points for monitoring student progress. Weak points in elementary programs could be detected and changed to assure student success, and testing at grades five and eight would provide a summary evaluation of the elementary and middle school program and articulation of student achievement to the middle and high school. The committee recommended that all students at these grade levels be administered these tests for the present, but that a sampling design should replace mass testing as criterion-referenced tests are developed.

To determine which specific standardized tests should be used for testing, curriculum department coordinators in reading and mathematics were first asked to evaluate the major published standardized reading and mathematics achievement tests for content validity and congruence with Madison educational objectives and submit the 2-4 most appropriate tests to the committee. A subcommittee from the Nucleus Testing Committee then evaluated the psychometric qualities of these tests. Where there was no detectable difference between the quality of a test already being used and other tests, the subcommittee recommended keeping the former to facilitate longitudinal comparisons and reduce costs.

Recognizing that one purpose of the Nucleus Testing Committee was to improve the use of test results, committee members worked in grade level groups to determine how test results should be reported to teachers and administrators to provide each user group with meaningful information. In accordance with committee recommendations, reports were distributed to the following audiences:

Teachers

(1) A frequency distribution by class (both national and local norms)

(2) Class lists, including student name, raw score, and norms

(3) Item analysis (optional)--requested by 15% of the schools

(4) Test subscore results, e.g., vocabulary and comprehension scores for the STEP Reading (Optional)--requested by 15% of the schools

Principals

(1) A school-wide frequency distribution with both national and local norms

All Administrators

(1) Tables of equivalent rational and local norms

(2) Tables of system-wide frequency distributions

(3) Tables summarizing item-analysis data (lower 27%; upper 27%)

(4) Tables analyzing results by school and attendance area within local quartile

(5) Informal verbal reports based on the preceding tables

Several significant aspects of the standardized testing program outlined here should be noted: (1) group intelligence tests were eliminated, (2) one achievement test rather than a battery was recommended except at the transition between elementary and middle school, (3) standardized testing was viewed as one component of a testing program which also includes criterion-referenced and affective test development and (4) a sampling design was proposed for obtaining norm-referenced data once criterion-referenced measures are developed.

Response to Recommendations

An important step in the process was communicating with the subsystems that would be affected by changes in a testing program. Teachers and administrators received regular information through a bi-weekly newsletter. The set of preliminary recommendations for the testing program was presented to a city-wide parent advisory group which responded favorably to the elimination of IQ testing, reduction of time spent in testing, and the overall attitude of the committee toward the use of standardized tests to measure the success of the school system as a whole rather than the individual student. And it was to the credit of the members of the Nucleus Testing Committee and its leadership that the Instructional Cabinet, made up of the Superintendent, the Assistant Superintendent and the instructional directors, approved the recommendations of the committee for the city-wide testing program as presented.

Implementation and Evaluation

The implementation of the standardized testing program has been carried out under the direction of the Coordinator of Research and Testing. Nucleus Testing Committee members in the elementary and middle schools assisted in the test administration phase by conducting an orientation session for teachers in their buildings on how the tests were to be administered and scored. A Teacher Quality Control checklist on testing procedures was completed by teachers. On the whole, data received from the schools was in good order, which allowed for better turnaround on results and analysis and, more importantly, for greater confidence in the test results.

Computer services within the school system were utilized for analysis of the data. Test reports were generated as recommended by the Nucleus Testing Committee. In addition, informal verbal reports addressing specific questions that could be asked of the data on a city-wide basis have been prepared by the Research and Testing Office and distributed to principals, central office administrators and committee members.

One of the goals set by the Nucleus Testing Committee for itself at the beginning of the present school year was to evaluate the revised standardized testing program. Three major questions were addressed to the test users, i.e., randomly selected teachers, specialized education staff, principals, and central office administrators, and the Coordinator of Research and Testing, through questionnaires developed by Committee members: (1) What use have you made of the data provided

to you in the standardized test reports? (2) What problems were encountered in test administration and data analysis? and (3) Do you have need for standardized test data beyond the revised city-wide program and if so, for what purpose? Revisions for the 1972-1973 city-wide testing program will be made by the committee on the basis of this evaluation.

No less important is the cost-benefit factor of the revised program. Three years ago it was estimated that testing cost the school system about $80,000 a year, including teacher time, administrator time, secretarial time, computer processing time, and testing materials. It is estimated that the expense was reduced to about $35,000 for all tests given this year.

Looking to next year, the Coordinator of Research and Testing is presently working with our Management Information System in developing a new computer-assisted test scoring package that will offer greater flexibility in test reports. The user will have more control over the type of information he receives from testing by selecting the kind of reports he needs to make educational decisions. Options in test result summaries will include percentile bands and stamines, summary statistics, and test reliability statistics. Names of students who fall below a criterion raw score (determined by user) will be marked with an asterisk for quick identification on the class list if requested. An item-pupil response matrix will be available to cluster items by the concept measured and indicate correct, incorrect, or no response for each item, allowing for greater interpretation of test results.

Demographic information selected by the user will also be reported to further examine patterns in achievement.

Summary

The city-wide standardized testing program of Madison Public Schools was reviewed by a committee of a cross section of school system educators as part of a total effort to design a testing program more sensitive to the needs of the system. Based on an examination of high intercorrelations among subtests of major standardized tests and results of a survey of Madison educators indicating little need for standardized test results beyond reading and mathematics, standardized testing was reduced to reading (grades 1, 2, 3, 4, 5, and 8) and mathematics (grade 5). Levels of administration were determined by the importance of measuring reading progress in elementary grades and the value of achievement level indicators in transition between elementary, middle and high school. Standardized tests, under this plan, are intended to provide normative data to compare our school system with others in the nation, to evaluate educational programs within the system, and to give an indication of student ranking in achievement to teachers until more specific criterion-referenced tests are adopted. Testing of every student will then be replaced by a random sampling design that will provide the same normative data for less time and cost.

The program outlined has been implemented during the present school year with involvement of committee members in the quality control of data obtained through testing. Separate test reports and interpretation issued to teachers and administrators have increased use of the data. Evaluation of the entire program is showing areas for further refinement in computer analysis and interpretation to assist the consumers in utilizing the test data for making more substantive educational decisions.

REFERENCES

1. Gates, Arthur I., and Walter H. MacGinitie. Gates-MacGinitie Reading Tests: Technical Manual, New York: Teachers College Press, 1965.

2. Lindquist, E. F. and A. N. Hieronymus. Iowa Tests of Basic Skills: Manual for Administrators, Supervisors and Counselors, New York: Houghton Mifflin Company, 1964.